MetaGrey Europe, A Proposal in the Aftermath of EAGLE-SIGLE

Schöpfel, J.

Bionote: Joachim Schöpfel graduated from the University of Hamburg in 1984. A research assistant and lecturer at the University of Hamburg, Department of Developmental and Educational Psychology, from 1985 to 1990, he obtained his Ph.D. from the same university in 1992. He is head of the e-publishing and document supply department at INIST-CNRS and teaches Culture and Society (1992-2001) and Documentation (from 2001 on) at the University of Nancy. He is member of the UK Serials Group and last president of EAGLE.

Contact address: Dr. Joachim Schöpfel, INIST-CNRS, 2 allée du Parc de Brabois, 54514 Vandoeuvre Cedex, France. E-Mail schopfel@inist.fr

Abstract

In the aftermath of EAGLE, the objective of the MetaGrey Europe project is to organize a new network of European leading organisations for the collection and dissemination of grey research literature, to facilitate the access to European grey literature through a unique web-interface (metasearch engine), to increase its international visibility and to safeguard the former SIGLE records and integrate them in the new network in an OAI-PMH compliant format. The communication describes the project scope, assesses the resources, contains a risk analysis and an outline of costs and benefits and provides information about project management, communication strategy and planning.

Introduction

2005 is a crucial year for the European grey literature*. The EAGLE* concept doesn't work any longer. The input to the SIGLE* database definitely stopped in April, and the EAGLE association is to be liquidated.

EAGLE was a co-operative network for identification, location and supply of grey research literature. A non-profit association formed by the national centres participating in SIGLE, it produced and provided access to the bibliographic database SIGLE. All documents referred to could (and still can) be obtained through the national SIGLE centres (see Wood & Smith 1993).

The decline of the EAGLE concept has four reasons:

(1) Internet: The 1985 concept of the SIGLE database has not been able to keep up with the rapid development with the Internet and the online resources. Even if FIZ Karlsruhe* developed a web-version of SIGLE, the database production didn't adjust to the possibilities offered by the new technologies of information:

- No direct online cataloguing in a shared database,
- no metadata harvesting,
- very few records of e-documents[1],

- few direct access to full text,
- few links to other resources.

(2) Organisational structure: The EAGLE structure (national input centres, a central operating agent) was unable to open the network easily to new and/or important producers of grey literature on the national or international level. Furthermore, it was not adapted to the need of flexible and rapid decision taking and marketing.
(3) Coverage: For most of the EAGLE countries, the national input became less and less representative. Often, even for the national input centres input criteria and workload excluded a significant part of their grey collections (dissertations, reports, conference proceedings)[2].
(4) Economic model: The EAGLE model was based on flat membership fees and revenues from servers. Access to the database remained "traditional" with an annual subscription to the host or else a pay-per-view for each record with a credit card. With important members retiring from input and membership and the success of open access based initiatives, the economic model came to an end and needed revision. Investment for the development of the database was not provided.

The EAGLE members started a strategic debate on the future of the concept in 2001. The result is the present project: instead of a further development of the 20-year-old concept, the German and French EAGLE members propose a radical change that integrates the latest web-based technologies and the economic environment of open access to scientific and technical literature (for the open access trends in Europe and France, see Grüttemeier & Mahon 2003 and Grüttemeier 2005).

The following communication is an updated and augmented version of the project brief discussed at the 2005 General Assembly of EAGLE in Karlsruhe. Acknowledgments to Silke Rehme (FIZ Karlsruhe) for her helpful comments.

Projects aims and objectives

The main objective of the MetaGrey Europe project is to translate the key ideas of EAGLE in terms of 21st century information technology:

1. To organize a network of European leading organisations for the collection and dissemination of grey research literature.
2. To facilitate searchability of European grey literature through a unique web-interface.
3. To increase its international visibility.
4. To enhance online access to the full-text of European research grey literature whenever possible.
5. A complementary objective is to safeguard the SIGLE records (file data: 856,610 records) and integrate them in the new network.

The target group of MetaGrey Europe are scientists, academics, students, and library and information professionals.

Project scope

The MetaGrey Europe project's objectives are declined into three rationales:

1. *Development of a metasearch engine for grey literature ("MetaGrey").*

The MetaGrey search engine is hosted on a central server and maintained by a designed member of the network. It provides a simple and intuitive search interface adapted to the bibliographic and metadata of the main types of grey literature. The metasearch engine technology developed since about 10 years has arrived at a stage of maturity that allows specific developments as for this project[3].

2. *Transformation of the SIGLE database in an open archive ("OpenSIGLE") hosted on an open archive server and OAI-PMH* compliant.*

The OpenSIGLE database is hosted on an OAI server and maintained by a designed member of the new network.

3. *Connexion of the most representative and important collections of European grey literature and of the "OpenSIGLE" archive to this metasearch engine.*

Grey literature key collections (databases, catalogues, archives) and other search engines are connected to the MetaGrey search engine by their institutions, with assistance by one or more members of the new network. There is no need for secondary or complementary cataloguing as for the SIGLE records.

The quality of the search engine and the overall projects depends also on the careful selection of important grey collections and reliable partner organisations that guarantee the quality of data and documents. The criteria of selection need thorough discussion. Special attention should be paid to the grey collections of the National and State Libraries. Conditions of access to and dissemination of grey literature are maintained by and remain under responsibility of the participating institutions. The overall structure of MetaGrey is decentralized.

The project rationale can be compared to the GrayLit portal of US Federal Agencies' grey literature.

Other projects in the MetaGrey environment are:

- The gateway "Science.gov" to authoritative selected science information provided by U.S. Government agencies, including research and development results.
- The new Google portal "Scholar" (access to scholarly literature from academic publishers, learned societies, universities and institutional repositories).
- The technology and concepts developed by the Hannover Search Engine Lab (with the leading German metasearch engine "MetaGer.de" and the German ResearchPortal "forschungsportal.net").
  http://meta.rrzn.uni-hannover.de
- The metasearch engine "Virtueller Katalog Kunstgeschichte" funded by the DFG* and developed with the MPG* (History of Art, 14 organisations). http://www.ubka.uni-karlsruhe.de/vk_kunst.html
- Different open access initiatives: the LARA* project of INIST-CNRS (Stock et al. 2005), the former MAGiC* project of the British Library and Cranfield University (Needham et al. 2002), the CERN* server (Vesely et al. 2004), and the CCSD* archives (Charnay 2003).

Whenever possible, the European projects should be included in theMetaGrey project.

Stakeholders

The project distinguishes three levels of participants:

1. A small group of core stakeholders for the project management, the communication strategy, the initial development of the "MetaGrey" software and the "OpenSIGLE" archive and the connexion of the most important grey collections. Today, three institutions are part of this core group: the TIB Hannover, INIST-CNRS and to a lesser degree, the FIZ-Karlsruhe.

2. A limited group of European national and special libraries and documentation centres with important grey collections. The following institutions have already been contacted and have shown interest for a participation in this project: the British Library and the Cranfield University (United Kingdom), the Dutch Royal Library (Koninklijke Bibliotheek), the CNR* (Italy), the CERN (Switzerland), the VNTIC* (Russia), the State Technical Library and Academy of Sciences of the Czech Republic.

3. An unlimited number of potential participants with significant grey collections and databases. For instance, the Slovak Centre of Scientific and Technical Information, the Spanish CINDOC-CSIC*, the Portuguese Fundacao para a Ciencia e a Tecnologia (FCT) Serviço de Informação e Documentação or the French ABES*, CCSD-CNRS, and Cyberthèses*. Actually, contacts exist with organisations located in 16 EU countries. Special attention should be paid to CORDIS (EU research reports).

Resource assessment

A preliminary resource assessment of the preparation, development and realisation of the MetaGrey project identified four domains with specific funding, staff and technical needs:

(1) Development of a metasearch engine for grey literature ("MetaGrey"):

1. Definition of content (searchable data, classification schema) and connexion interfaces.
2. Development of the software (with access to document delivery services).
3. Implementation on a server.

(2) Transformation of the SIGLE database in an open archive ("OpenSIGLE") hosted on an open archive server and OAI-PMH compliant:

1. Downloading of the SIGLE records and transformation into a harvestable structure.
2. Uploading of the OpenSIGLE data on an OAI server.

(3) Connexion of collections of European grey literature and of the "OpenSIGLE" archive to this metasearch engine:

1. Selection of collections and institutions (limited group).
2. Connexion of databases and catalogues (limited group).
3. Connexion of other collections (unlimited number).

(4) Functioning of MetaGrey Europe:

1. User assistance (helpdesk, connexion of new collections).
2. Search engine maintenance.
3. OpenSIGLE maintenance.

Up to now, a more detailed funding estimation exists only for the 2nd domain, the transformation of the SIGLE database. The INIST-CNRS information systems department made the following proposal (cf. communication to all former EAGLE member of July 5, 2005):

"This project scope is the transformation of the existing SIGLE database into an open archive hosted on an open archive server and OAI-PMH compliant. The idea is to convert the SIGLE records following the international metadata standards (Dublin Core mapping) and to load these metadata on an open archive that would be freely available to all users, with metadata harvesting via the OAI-PMH.

The INIST proposal is to realize the open archive OpenSIGLE with the DSpace software, to convert the SIGLE records, to load these data in the open archive, and to host the archive on a local INIST server.

The IS departement of INIST estimates the costs and charges of this project as follows :

Development of the software (based on DSpace): 20 days.
Conversion of records: 5 days.
Design, customization: 5 days.
Tests, validation: 10 days.

We can't estimate today how much time it would take to upload 800.000+ records in such an environment. For financial reasons, we would upload the records on only one server, not on two.

The estimation of 40 days of development is a low estimation. The economic minimum for this proposal is 25.000 euros: 20.000 euros for the development (software) and 5.000 euros for the server (hardware). If the remaining funds left for this project are higher, we could use this sum for instance to install a 2nd server, or for a more elaborated development. Anyway, this would be discussed with those of the members who agree with this proposal."

The submission of a Specific Targeted Research Project (STREP) to the European Commission (6th EU Framework Programme for Research and Technology Development or 6th FP) for an initial funding in 2005 was abandoned because of the too short delay after the EAGLE GA (deadline March 2005).

Risk analysis

Initially, six potential risks for the project had been identified. *In italics, updated commentaries*:

(1) Initial investment: No initial funding by the European Commission. No or limited funding by the EAGLE association. Restricted funding by the leading institutions. *Effectively, no EU funding in 2005, no EAGLE funding because of the unwillingness of the former members to spend "their" money for this project. No engagement by the "core institutions" up to now. Probably, there will be no funding demand for the 6th FP, and we should prepare a proposal for the 7th FP (2007-2013).*

(2) Language: Impossibility to select collections and data with a core language (English). *No*

real analysis of this topic up to now. Probably, an important part of European grey collections (more than 50%?) are not available and searchable in English.

(3) Bibliographic data: Too high complexity of data definition of the different types of documents, records and collections. *This needs analysis; nevertheless, the former EAGLE members already faced this problem – successfully. So why not now?*

(4) Connexions: Technical difficulties and limited motivations of institutions of the "limited group". *May be a real problem for some organisations.*

(5) OpenSIGLE: Technical and/or legal problems (intellectual property of SIGLE records) with the transformation of the SIGLE database into an open archive. *All former EAGLE members declared last summer that "their" SIGLE records could be temporarily archived by the EAGLE operating agent, the FIZ Karlsruhe, to be used by a future non-for-profit OAI compliant database. The technical aspects seem relatively simple.*

(6) Liquidation of EAGLE: Delay. No solution for the intellectual property of the SIGLE records. *The delay is caused by the liquidation of the bank accounts and the suppression in the Luxemburg Register for Societies and Commerce. Apparently, there are no major legal problems.*

#	Potential risk	Likelihood	Impact on project	Evaluation	Contingency
1	*Initial investment*	High	Medium	Most important risk	Alternative funding? Self-funding?
2	*Language*	Medium	High	Important risk	Acceptance of multilingual data
3	*Data*	Low	High	Low risk	Reduction of complexity
4	*Connexion*	Low	High	Low risk	Communication, adding other interfaces?
5	*OpenSIGLE*	Low	Low	Low risk	No alternative
6	*EAGLE*	Low	Medium	Low risk	Legal problems?

Table 1: Risk analysis

To resume, two risks (# 5 and 6) no longer exist. Two other risks (#2 and 3) should be analysed more thoroughly, based on concrete data and not on abstract considerations; this may need some months and could be done while prepare the funding demand. The risk #4 should also be assessed during the initial phase of the project. The most important risk, the need for financial resources (risk #1) today has no solution. For different reasons, most of the former EAGLE members seem to be exposed to financial shortcuts and a "fragile" budget situation. Without funding, no development. But without project rationale and minimal consensus on what to do and with whom, no needs analysis and no funding demand. So, where to start?

Another risk needs evaluation: the need of ongoing and permanent funding of the metasearch engine (maintenance, developments, helpdesk).

Two other risks have been excluded: (a) The technology risk can be neglected because of the maturity of the used software concept. (b) Actually, there seems to be no concurrent project with analogue objectives or scope. Nevertheless, this situation can change very quickly. Elsevier announced recently (6 Octobre 2005) a "landmark partnership" with the Networked Digital Library of Theses and Dissertations (NDLTD) to add the extensive collection of theses and dissertations to the Elsevier free science-specific search engine, Scirus. This partnership includes indexing the NTLTD content on Scirus.com and powering the search service on the repository's site. If Scirus is to become a free search engine that indexes important grey collections, is there a real need (and funding) to another initiative on the same market sector?

Outline of costs and benefits

The economic model of MetaGrey Europe is a sustainable, low-budget project, based on limited funding of the initial phase (software development and server, marketing and communication strategy) and use of existing resources (staff and budget of member organisations) for the functioning of the metasearch engine (maintenance, assistance) and the OpenSIGLE database (maintenance).

Apart from the connexion to the search engine, the participating organisations will not assume any financial obligation. Technical input may be necessary only in case the partners change their own parameters (database format, cataloguing standards etc.).

The benefits for the participating organisations are twofold: the metasearch engine increases the visibility and searchability of their special grey collections through a new and dedicated "access point"; and it potentially increases the use of their document supply and ILL services through the link to their own online interfaces.

Project management approach

The proposed steering committee is composed of four organisations, each of them with specific roles:

INIST-CNRS (France): Project coordination, submission of EU funding project, definition of searchable data and core collections, user assistance. Development and host of OpenSIGLE.
TIB Hannover (Germany): Development, host and maintenance of MetaGrey. Definition of searchable data and core collections.
FIZ Karlsruhe (Germany): Liquidation of EAGLE. Temporary archiving of the SIGLE data.
GreyNet* (Netherlands): Development and realisation of marketing and communication strategy, promotion of the new network.

The cooperation between these organisations could be facilitated by a listserv with restricted access and hosted by one of these organisations.

Marketing and communication strategy

The marketing and communication strategy is developed and coordinated by GreyNet. The multilevel strategy includes several vectors of communication and information:

EAGLE network: Information and promotion through the existing mailing list of former and actual EAGLE member organisations.
CNRS network: Information and promotion through the recently created CNRS network of direct scientific communication organisations.
GL conferences: Communications at the GreyNet international conferences on grey literature 2005 (France) and 2006 (United States).
Other conferences: Information and promotion at selected national and international conferences and events on scientific information and open access.
Listservs: Information and promotion through selected national and international listservs of library and information professionals and scientists (ADBS-Info, Biblio.fr, INETBIB, JISC listservs, SPARC-Europe).
Publications: Information and communication in selected LIS serials (for instance, *Documentaliste*, *Bibliotheksdienst*, *Interlending and Document Supply*, *The Grey Journal, Electronic Library*, *Aslib Proceedings: New Information Perspectives*, *Library Hi Tech News*, *Online Information Review* etc.).

Timelines and milestones

Seven key stages had been identified. The timelines mainly depend on the submission, evaluation and negotiation of the initial funding demand. The following table contains the initial project agenda (February 2005):

#	Project phase	Content	Period	Commentary
1	Preparation (1)	Concept, project team, network	2004 November – 2005 January	
2	Preparation (2)	Elaboration and submission of a funding proposal	2005 January – 2005 March	Depends on FP6 calls for proposals
3	EAGLE	Liquidation of association	2005 January – 2005 June	
4	Development (1)	MetaGrey Europe	2005 April – 2005 December	Depends on #2
5	Development (2)	OpenSIGLE	2005 April – 2005 December	To be specified, depends also on #3
6	Connexion (1)	First group	2005 September – 2006 March	During #4
7	Connexion (2)	Unlimited number	2006 April – 2007 March	Following #4 and 6

Table 2: Timelines

One year after the first project draft, the initial timelines are of course no longer realistic and need to be updated. Nevertheless, the table shows clearly two "milestones": first, the need to find a minimal consensus on the concept (#1); second and based on this consensus, the need to assure an initial funding (#2).

Concluding remarks, open questions

At the end of 2005, we can observe small progress: the legal and financial liquidation of the European association is nearly finished, and FIZ Karlsruhe guarantees the temporary safeguard of the SIGLE records.

Unfortunately, once the SIGLE input stopped, technical and strategical coordination between the former EAGLE centres ceased, too. Our organisations seem preoccupied with their own problems, putting priorities elsewhere, trying to redefine their objectives and activities in the environment of the Open Access movement, the Google Print initiative and the EU intellectual property rights.

In some way, we are facing a similar situation as 20 years ago, before the creation of the EAGLE network. Will there be a second chance for European grey literature? And if so, when? Will the European projects for open institutional repositories and digital libraries attract too many financial, human and technical resources without leaving space for a new grey project? Do we need some more years to measure its real interest? Will there be a concurrent project in the same sector (e.g. Elsevier)? Maybe the Nancy conference on grey literature will provide a first answer.

Glossary

ABES: Agence Bibliographique de l'Enseignement Supérieur (France), a public agency that hosts the French national academic union catalogue.

CCSD-CNRS: Centre de Communication Scientifique Directe of the French CNRS with a number of open archives (preprints, electronic theses and dissertations, articles, proceedings).

CERN: European organisation for nuclear research (Switzerland).

CINDOC-CSIC: Centro de Informaçion y Documentaçion Cientifica, Centre of Scientific Information and Documentation of the Spanish Higher Council for Scientific Research.

CNRS: Centre National de la Recherche Scientifique, a public organisation for scientific and technological research under the authority of the French Ministry of Research (France).

CNR: Consiglio Nazionale delle Ricerche, public research organisation (Italy).

DFG: Deutsche Forschungsgemeinschaft, public research agency under the authority of the German Ministry of Research (Germany).

EAGLE: European Association for Grey Literature Exploitation (Luxemburg).

FIZ Karlsruhe: Fachinformationszentrum Karlsruhe, a non-profit scientific service institution that produces and markets scientific and technical information services (Germany).

Grey literature: "Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing" (GL'97 Luxemburg Convention).

GreyNet: Grey literature network service (Netherlands).

INIST-CNRS: Institut de l'Information Scientifique et Technique (France), the CNRS documentation centre for scientific and technical information.

LARA: Libre Accès aux Rapports Scientifiques et Techniques, an open archive of French reports under construction by INIST-CNRS (France).

MAGiC: A former UK project sponsored by the British Library and the Research Support Libraries Programme and hosted by the Cranfield University to enhance awareness, access and utilisation of key collections of technical reports.

MPG: Max-Planck Gesellschaft, a public organisation for scientific and technological research under the authority of the German Ministry of Research (Germany).

OAI-PMH: Open Archive Initiative-Protocol of Metadata Harvesting.

RRZN: Computer Centre of the Lower Saxony, University of Hannover (Germany).

SEL: Search Engine Laboratory of the University of Hannover (Germany).

SIGLE: System of Information for Grey Literature in Europe.

TIB Hannover: Technische und Informationsbibliothek Hannover, most important German library for technical and scientific information.

VNTIC: Scientific and Technical Information Centre (Russia).

References

Charnay, Daniel: "The Centre for Direct Scientific Communication". In: Grüttemeier & Mahon 2003, 133-137.

Grüttemeier, Herbert, & Mahon, Barry (ed.): "Open Access to Scientific and Technical Information. State of Art and Future Trends". ICSTI/INIST/INSERM Seminar, 23-24 January 2003, Paris, France. Amsterdam: IOS Press 2003.

Grüttemeier, Herbert: "The Way to Open Access – French Strategies to Move Forward". *Library and Information Service (Tushu Qingbao Gongzuo)* 2005 (to be published).

Needham, Paul A.S., Sidwell, Katy, Bevan, Simon J., & Harrington, John: *The MAGiC Final Report.* London: British Library Cooperation and Partnership Programme 2002. http://hdl.handle.net/1826/58

Stock, Christiane, Rocklin, Emmanuelle, & Cordier, Aurélie: "LARA: Open Access to Scientific and Technical Reports". *GL7 Conference. Seventh International Conference on Grey Literature: Open Access to Grey Resources.* Nancy, 5-6 December 2005.

Vesely, Martin, Baron, Thomas, Le Meur, Jean-Yves, & Simko, Tibor: "CERN document server: Document management system for Grey Litetarture in a networked environment". *Publishing Research Quaterly*, 2004, 20, 1, 77-83.

Wood, David N., & Smith, Andrew W.: "SIGLE: a model for international co-operation". *Interlending & Document Supply*, 1993, 21,1 18-22.

[1] We could identify with certainty only 578 documents on CD-ROM (0.07%). It is impossible to determine by searching in the database the number exact of online resources.

[2] Here the table with the coverage by member countries of the SIGLE database from February 2005 (total 855,260 records) :

Country	Records	In %
Belgium	5,236	0,6
Czech Republic	5,683	0,7
France	72,752	8,5
Germany	198,588	23,2

Hungary	728	0,08
Ireland	97	0,01
Italy	30,758	3,6
Latvia	3,916	0,45
Luxemburg	33	0,0
Netherlands	61,471	7,2
Poland	87	0,01
Portugal	1,780	0,2
Russian Federation	19,524	2,3
Slovakia	800	0,09
Spain	11,256	1,3
Sweden	14	0,0
United Kingdom	434,007	50,7
Commission of the EC	8,530	1,0

The French example may illustrate the non representativity of the national input: the INIST library contains about 150,000 theses, 70,000 conference proceedings, and 70,000 scientific and technical reports - more than four times the number of French SIGLE records. Even if all these documents are not grey, most of them are. Other important "grey" producers never contributed to the national input (for instance, the national Atomic Energy and Space Research Centres).

[3]See for instance the developments and publications of the Hannover University Search Engine Lab (W. Sander-Beuermann) at http://metager.de/suma-eng.html